Generating functions of circular codes

Frédérique Bassino
I.G.M., Université de Marne La Vallée
2, rue de la butte verte 93166 Noisy le Grand Cedex
e-mail: bassino@univ-mlv.fr
`http://www-igm.univ-mlv.fr/bassino`

Abstract

We describe in terms of different parameters the generating series of the star of a circular code. We extend the characterization of length distributions of circular codes established for a finite alphabet by Schützenberger to an arbitrary "weighted" alphabet. In this framework, we give a new characterization of these length distributions. This one directly concerns the coefficients of the generating series of the code instead of the number of primitive conjugacy classes. This result shows that we can decide whether a finite sequence is the length distribution of a circular code. We also establish a necessary and sufficient condition for a series to be the length distribution of a maximal circular code over a finite alphabet.

Introduction

Codes are the basis of free submonoids, *i.e.*, are the sets of words with the property of unique decipherability. Circular codes are a special class of codes. They are the sets of words with the property of unique decipherability on a word written on a circle.

They were introduced by Golomb and Gordon ([GG65]) under the name of codes with bounded synchronization delay, because they have a strong property of synchronization. Thus any thin circular code is synchronous and any uniformly synchronous code is circular. For this reason, they play an important role in problems of error correction.

A particular kind of circular codes, the coma-free codes ([GGW58]) have been the topic of many investigations. Circular codes have many applications, in particular, in the analysis of DNA sequences. Recently, by a study of the repartition of trinucleotides in protein coding genes according to the reading frame, D. Arquès and C. Michel have identified three sets of codons constituting circular codes ([AM95]). They can be obtained, one from the other, by a circular permutation.

Circular codes also appear in many problems of combinatorics of words. For example, in order to compute the number of circular words on the alphabet $A = \{a, b\}$ that do not have two successive occurrences of the letter b, one can use the circular coding morphism $\alpha : A^* \to A^*$, defined by:

$$\alpha : \begin{array}{l} a \to a \\ b \to ba. \end{array}$$

The coefficient f_n of z^n in the generating series of the submonoid $\alpha(A^*)$ generated by the circular code $\{a, ba\}$

$$\sum_{n \geq 0} f_n z^n = \frac{1}{1 - z - z^2}$$

is then equal to the number of circular words of length n ($n \geq 2$), not having bb as factor. An analogous result holds for any finite circular code. Other examples of combinatorial problems solved with the help of circular codes are presented in [Sta86] (Section 4.7). Finally circular codes are linked with finite factorizations of the free monoid: all factors occurring in such a factorization are circular ([Sch65a], [Vie78], [Kro87]).

In this paper, we prove several new results. We describe in terms of different parameters the generating series of the star of a circular code. We extend the characterization of length distributions of circular codes established for a finite alphabet by Schützenberger to an arbitrary "weighted" alphabet. We give a new form of it, that directly concerns the coefficients of the generating series of the code instead of the number of primitive conjugacy classes. This result shows that we can decide whether a finite sequence is the length distribution of a circular code. Finally, we establish a necessary and sufficient condition for a sequence of nonnegative integers to be the length distribution of a maximal circular code over a finite alphabet.

In Section 1, we recall some definitions and results related to the theory of codes ([BP85]) and particularly to the conjugacy relation. We also introduce the notion of weighted alphabet, which is nothing else but a set of symbols to which integral weights are associated. Shannon ([Sha48]) was one of the first to be interested in alphabets whose letters have arbitrary costs, that, in this case, represent the duration of transmission of the symbols. These alphabets were subsequently used in information theory to generalize the notion of entropy that he proposed ([Csi69], [Kra62]). Depending on the applications, the costs are parameters linked with the magnitude, the duration or the energy of a signal.

In next section, we give several formulas for the generating series of the star of a circular code and establish among others variants of Witt formula and cyclotomic identity.

In Section 3, we extend (Theorem 2) the characterization of length distributions of circular codes established for a finite alphabet by Schützenberger ([Sch65b]) to an arbitrary "weighted" alphabet and obtain by this way a result concerning two length distributions (Corollary 2).

In Section 4, we obtain, by combinatorial methods, a new characterization (Theorem 3) of length distributions of circular codes, that underscores the decidability in the case of a finite distribution.

The notion of maximality of codes (see survey [BL96]) occurs in numerous problems linked to the optimization of a coding procedure. In a certain way, maximal codes are those that make use of the whole capacity of the transmission channel. We give, in the last section, a necessary and sufficient condition for a series with nonnegative integral coefficients to be the length distribution of a maximal circular code over a k-letter alphabet (Theorem 4).

1 Preliminary

Weighted alphabet

We define a *weighted alphabet* as a pair (A, λ) consisting in an alphabet A and in a morphism $\lambda : A^* \to \mathbb{N}$ defined by

$$\begin{cases} \lambda(1) = 0, \quad \text{where 1 is the empty word,} \\ \forall a \in A, \quad \lambda(a) \in \mathbb{N}^*, \\ \forall u, v \in A^*, \quad \lambda(uv) = \lambda(u) + \lambda(v), \end{cases}$$

i.e., for all $w \in A^*$, $\lambda(w) = \sum_{a \in A} \lambda(a)|w|_a$, where $|w|_a$ denotes the number of occurrences of a in w. The integer $\lambda(w)$ is then called the *length* of the word w of (A^*, λ) . The morphism λ can be interpreted as a cost as well as a distance or length.

Remark 1 A "usual" alphabet A can be seen as a weighted alphabet for which the morphism λ is identically 1 over A. In this particular case, we will denote by $|w|$ instead of $\lambda(w)$ the length of a word w.

The definitions and results established for a "usual" alphabet, for which the notion of length does not occur, stay the same for a weighted alphabet.

Codes

In the following, we recall some definitions and results related to the theory of codes (see [BP85]).

A subset X of the semigroup A^+ is a *code* over A if and only if

$$\forall n, m \geq 1 \quad \text{and} \quad x_1, \ldots, x_n, x_1', \ldots, x_m' \in X$$

$$x_1 x_2 \cdots x_n = x_1' x_2' \cdots x_m' \quad \Rightarrow \quad n = m \quad \text{and} \quad x_i = x_i' \quad \text{for} \quad i = 1, \ldots, n.$$

In other words, a set X is a code if any word of X^+ has a unique factorization into words from X.

A subset $X \subset A^+$ is a *circular code* if for all integers $n, m \geq 1$ and $x_1, x_2, \ldots, x_n \in X$, $y_1, y_2, \ldots, y_m \in X$ and $p \in A^*$ and $s \in A^+$ the equalities

$$s x_2 x_3 \cdots x_n p = y_1 y_2 \cdots y_m \quad \text{and} \quad x_1 = ps$$

imply

$$n = m, \quad p = 1 \quad \text{and} \quad x_i = y_i \quad (1 \leq i \leq n).$$

In other words, any word of X^+ written on a circle (Figure 1) has a unique factorization.

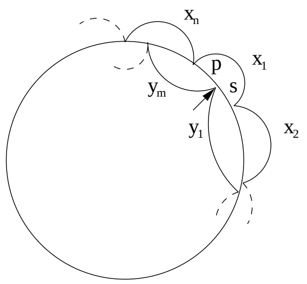

Figure 1: two circular factorizations

Note that a circular code is a code and any subset of a circular code is also a circular code.

Conjugacy

We now introduce the notion of conjugacy in a monoid, notion for which we also give some properties. This equivalence relation allows us to characterize circular codes.

Two words x and y of A^* are *conjugate* if there exist two words u and v of A^* such that $x = uv$ and $y = vu$. The classes of this equivalence relation are called *conjugacy classes* (and sometimes *circular words* or necklaces, see [Reu88] and [MR83]).

A word $x \in A^*$ is *primitive* if it is not a power of another word of A^* : x is primitive if, and only if, $(x = y^n \quad \text{and} \quad n \geq 0) \quad \Rightarrow \quad x = y$. The empty word is not primitive.

Remark 2 A circular code X cannot contain two distinct conjugate words. Moreover, all words of X are primitive.

Let X be a code. Two words w and w' of X^* are called *X-conjugate* if there exist $x, y \in X^*$ such that $w = xy$ and $w' = yx$. We call X-conjugacy classes the classes of this equivalence relation. A word $x \in X^*$ is called *X-primitive* if it is not a power of another word of X^*.

Remark 3 As the alphabet A is itself a circular code, the conjugacy relation can be seen as the relation of A-conjugacy and the notion of primitivity as the notion of A-primitivity.

Two words of X^* which are X-conjugate are conjugate. Likewise, a word in X^* which is primitive is also X-primitive. When X is a circular code, the converse of these two properties is true.

We now give a characterization of circular codes based on the conjugacy relation (see [BP85]). Let $X \subset A^+$ be a code, X is a circular code if, and only if,

$$\forall u, v \in A^*, \quad uv, vu \in X^* \qquad \Rightarrow \qquad u, v \in X^*.$$

Example 1 Let $A = \{a, b\}$ and $X = a^*b$, then $X^* = A^*b \cup \{1\}$. Thus if uv and vu are in X^*, each of the words u and v is either the empty word or terminates by the letter b, hence it belongs to X^*. Thus that X is a circular code.

Let $A = \{a_i / i \geq 0\}$ and $X = \{a_i a_{i+1} / i \geq 0\}$. One can easily verify that X is again a circular code.

Let $A = \{a, b, c, d\}$ and $X = \{ab, cd, bc, da\}$, the words $abcd$ and $bcda$ are in X^*, but neither a, nor bcd are in the code, so X is not circular.

2 Very pure submonoids

Let (A, λ) be a weighted alphabet and X be a subset of A^*. For all nonnegative integers n, we denote by A_n the set of words of length n in A^*:

$$A_n = \{w \in A^* / \lambda(w) = n\},$$

by α_n the number of words of length n in X:

$$\alpha_n = \mathrm{Card}\{x \in X / \lambda(x) = n\} = \mathrm{Card}\{x \in X \cap A_n\}$$

and by f_X the generating series of X defined by: $f_X(z) = \sum_{n\geq 0} \alpha_n z^n$. If A is the "usual" alphabet, the integer α_n is then the number of words of X having n letters. In a more general way, the sequence $(\alpha_n)_{n\geq 1}$ is called the *length distribution* of X.

Given a code X over a weighted alphabet A, we denote by $L_n(X)$ the number of X-conjugacy classes of X-primitive words in $X^* \cap A_n$.

Proposition 1 *Let A be a weighted alphabet and let $X \subset A^+$ be a circular code over A. Then,*

$$\forall n \geq 1, \qquad L_n(X) \leq L_n(A),$$

and the equality holds if, and only if, each conjugacy class of primitive words of A_n meets X^*.

This result is a generalization of a result known in the case of circular codes over a usual alphabet (see [BP85] Proposition 3.2 p.338) and is obtained by replacing the usual notion of length by its extended version.

Moreover, from a result due to Golomb and Gordon ([GG65]), when X is a code over a k-letter alphabet, the numbers $L_n(X)$ depend only upon the length distribution of the code X. This property still holds when X is a code over an arbitrary weighted alphabet.

Indeed, let $\alpha = (\alpha_n)_{n\geq 1}$ be the length distribution of X. We denote by $\alpha_n^{(i)}$ the number of words of length n in X^i ; as X is a code, one has $f_{X^i}(z) = (f_X(z))^i$, thus $\alpha_n^{(i)}$ is the nth coefficient of the series $(f_X(z))^i$. We set

$$\forall n \geq 1 \qquad s_n(\alpha) = \sum_{i=1}^{n} \frac{n}{i} \alpha_n^{(i)}.$$

Note that, when X is a circular code, the integer $s_n(\alpha)$ can be interpreted as the number of words of length n whose conjugate is in X^*. We then get

$$\forall n \geq 1, \qquad L_n(X) = l_n(\alpha) \quad \text{with} \quad l_n(\alpha) = \frac{1}{n} \sum_{d/n} \mu(\frac{n}{d}) s_d(\alpha),$$

where μ denotes the Möbius function.

Remark 4 As the alphabet A is itself a circular code, the numbers $L_n(A)$ depend only upon the length distribution of the alphabet, and thus in the particular case of a finite "usual" alphabet, upon its cardinality k.

A submonoid N of a monoid M is said to be *very pure* if $uv, vu \in N \quad \Rightarrow u, v \in N$. It can be shown that a submonoid of A^* is very pure if, and only if, its minimal set of generators is a circular code ([BP85] Proposition 1.1 p.323). We now describe in terms of different parameters the generating series of very pure submonoid of A^*.

Let X be a circular code over a weighted alphabet A and let $f_X(z) = \sum_{n\geq 1} \alpha_n z^n$ be its generating series. First, note that, since X is a code, the generating series f_X, f_{X^i} and f_{X^*} of X, X^i and X^* respectively, are linked by the relations (see [BP85])

$$f_{X^*}(z) = 1 + \sum_{i\geq 1} f_{X^i}(z) = \frac{1}{1 - f_X(z)}.$$

Proposition 2 *The generating series $f_{X^*}(z)$ of the very submonoid X^* generated by the circular code X satisfies*

$$f_{X^*}(z) = \exp\left(\sum_{n\geq 1} \frac{s_n(\alpha)}{n} z^n\right),$$

where $s_n(\alpha)$ is the number of words of length n whose conjugate is in X^*.

Proof : By definition,

$$\forall n \geq 1, \quad s_n(\alpha) = \sum_{i=1}^{n} \frac{n}{i} \alpha_n^{(i)} \qquad \text{and} \qquad \forall i \geq 1 \quad f_{X^i}(z) = \sum_{n\geq 1} \alpha_n^{(i)} z^n,$$

thus we get

$$\sum_{n\geq 1} \frac{s_n}{n} z^n \quad = \quad \sum_{i\geq 1} \frac{1}{i} f_{X^i}(z). \tag{1}$$

Moreover, as

$$\sum_{i\geq 1} \frac{1}{i} f_{X^i}(z) = \log\left(\frac{1}{1 - f_X(z)}\right),$$

taking the exponential of each member of the equality (1), we obtain

$$\frac{1}{1 - f_X(z)} = \exp\left(\sum_{n\geq 1} \frac{s_n(\alpha)}{n} z^n\right),$$

concluding the proof. □

Proposition 3 *The generating series $f_{X^*}(z)$ of the submonoid X^* generated by the code X also satisfies*

$$f_{X^*}(z) = \frac{1}{\prod_{n\geq 1}(1 - z^n)^{l_n(\alpha)}},$$

where $l_n(\alpha)$ is the number of X-conjugacy classes of X-primitive words of $X^* \cap A_n$.

Proof : From the cyclotomic identity, if A is a k-letter (usual) alphabet, then

$$\frac{1}{1 - kz} = \frac{1}{\prod_{n\geq 1}(1 - z^n)^{l_n(k)}},$$

where $l_n(k)$ is the number of conjugacy classes of primitive words of length n in A^*. This result can be extented to codes in:

$$\frac{1}{1 - f_X(z)} = \frac{1}{\prod_{n\geq 1}(1 - z^n)^{l_n(\alpha)}}. \tag{2}$$

Indeed, from Equality (1) and the relation $s_n(\alpha) = \sum_{d/n} d l_d(\alpha)$, we deduce

$$\sum_{i\geq 1} \frac{1}{i} f_{X^i}(z) = \sum_{n\geq 1} \sum_{d/n} d l_d(\alpha) \frac{z^n}{n}, \tag{3}$$

The last member of Equality (3) can also be written as

$$\sum_{n\geq 1}\sum_{d/n} dl_d(\alpha)\frac{z^n}{n} = \sum_{d\geq 1} l_d(\alpha)\sum_{q\geq 1}\frac{z^{qd}}{q} = \sum_{d\geq 1} l_d(\alpha)\log\left(\frac{1}{1-z^d}\right).$$

Taking the exponential of each member of Equality (3), we get (2). □

Proposition 4 *The generating series $f_{X^*}(z)$ of the submonoid X^* generated by the code X also satisfies*

$$f_{X^*}(z) = \frac{1}{\prod_{n\geq 1}(1-a_n(\alpha)z^n)} \qquad \textit{with} \quad f_X(z) = \sum_{i\geq 1}\alpha_i z^i,$$

where, for each n, $a_n(\alpha)$ is a polynomial with nonnegative integral coefficients in the variables $(\alpha_i)_{i\leq n}$. Moreover this sequence of polynomials can be computed recursively by making use of the following formula

$$\forall n \geq 1, \qquad a_n = \alpha_n + \sum_{i=1}^{\lceil \frac{n}{2}-1\rceil} a_{n-i}a_i,$$

where $\lceil x\rceil = \min\{n/n \geq x, n\in \mathbb{N}\}$.

Proof : A result of Gaudier ([Gau96]) shows that, for each n, $a_n(\alpha)$ is a polynomial with nonnegative integral coefficients in the variables $(\alpha_i)_{i\leq n}$. Indeed we prove here a more precise statement :

$$\forall n \geq 1, \qquad a_n = \alpha_n + \sum_{i=1}^{\lceil \frac{n}{2}-1\rceil} a_{n-i}a_i, \tag{4}$$

i.e., the first polynomials a_i are

$$a_1 = \alpha_1, \qquad a_2 = \alpha_2, \qquad a_3 = \alpha_3 + a_2a_1 = \alpha_3 + \alpha_2\alpha_1$$
$$a_4 = \alpha_4 + a_3a_1 = \alpha_4 + \alpha_3\alpha_1 + \alpha_2\alpha_1^2.$$
$$a_5 = \alpha_5 + a_4a_1 + a_3a_2 = \alpha_5 + \alpha_4\alpha_1 + \alpha_3\alpha_2 + \alpha_3\alpha_1^2 + \alpha_2^2\alpha_1 + \alpha_2\alpha_1^3.$$

In order to establish (4), we consider the sequence $(f_n(z))_{n\geq 1}$ defined by

$$\forall n \geq 1, \quad f_n(z) = \prod_{i\geq n}(1-a_i(\alpha)z^i),$$

the polynomial a_n is then equal to the opposite of the coefficient of z^n in $f_n(z)$. We have

$$f_1(z) = 1 - f_X(z) = 1 - \sum_{i\geq 1}\alpha_i z^i \qquad\qquad \forall n\geq 1, \quad f_{n+1}(z) = \frac{f_n(z)}{1-a_nz^n}.$$

We shall prove by induction on n that

$$\forall n \geq 1, \quad f_{n+1}(z) = 1 - \frac{\sum_{k=n}^{2n-1}(\sum_{i=1}^{\lceil \frac{k}{2}-1\rceil} a_{k-i}a_i)z^k}{\prod_{i\leq n}(1-a_iz^i)} - \frac{\sum_{i\geq n+1}\alpha_i z^i}{\prod_{i\leq n}(1-a_iz^i)}, \tag{5}$$

showing by this way (4). As $a_1 = \alpha_1$, we get

$$f_2(z) = \frac{1 - \sum_{i\geq 1} \alpha_i z^i}{1 - a_1 z} = 1 - \frac{\sum_{i\geq 2} \alpha_i z^i}{1 - a_1 z}.$$

Furthermore if the series $f_n(z)$ satisfies (5), a_n being equal to the opposite of the coefficient of z^n in $f_n(z)$, we have

$$a_n = \alpha_n + \sum_{i=1}^{\lceil \frac{n}{2} - 1 \rceil} a_{n-i} a_i.$$

Moreover as $f_{n+1}(z) = f_n(z)/(1 - a_n z^n)$, we obtain

$$f_{n+1}(z) = F - \frac{\sum_{k=n+1}^{2n-3} (\sum_{i=1}^{\lceil \frac{k}{2} - 1 \rceil} a_{k-i} a_i) z^k}{\prod_{i\leq n}(1 - a_i z^i)} - \frac{\sum_{i\geq n+1} \alpha_i z^i}{\prod_{i\leq n}(1 - a_i z^i)}$$

with

$$F = 1 - \frac{a_n z^n}{\prod_{i\leq n}(1 - a_i z^i)} = 1 - \frac{\sum_{i=1}^{n-1} a_n a_i z^{n+i}}{\prod_{i\leq n}(1 - a_i z^i)}.$$

Thus replacing F by this expression, we get (5), concluding the proof. □

3 Extension of a theorem of Schützenberger

Length distributions $(\alpha_n)_{n\geq 1}$ of codes over a k-letter alphabet are characterized by the Kraft-McMillan inequality: $\sum_{n\geq 1} \alpha_n k^{-n} \leq 1$. Among these sequences, Schützenberger identified with stronger conditions (Theorem 1) those that are length distributions of circular codes. In the following, we extend this characterization. This generalization concerns two points. On one hand, we replace the finite alphabet by an arbitrary weighted alphabet that allows us to extend the result to the case of two length distributions (Corollary 2). On the other hand, we reduce the set of values for which the inequalities have to be satisfied. This allows us to establish the decidability in the case of a finite length distribution (Corollary 1).

First, we recall the statement of the Schützenberger's result.

Theorem 1 (Schützenberger) *Let A be a usual k-letter alphabet. A sequence of nonnegative integers $(\alpha_n)_{n\geq 1}$ is the length distribution of a circular code over A if, and only if, it satisfies*

$$\forall n \geq 1, \qquad l_n(\alpha) \leq l_n(k)$$

where $l_n(k)$ and $l_n(\alpha)$ are the number of conjugacy classes of primitive words of length n in A^* and in X^* respectively.

Schützenberger's theorem can be extended as follows.

Theorem 2 *Let (A, λ) be a weighted alphabet and $\kappa = (\kappa_n)_{n\geq 1}$ be its length distribution. Then a sequence of nonnegative integers $\alpha = (\alpha_n)_{n\geq 1}$ is the length distribution of a circular code $Y \subset A^+$ if, and only if,*

$$\forall n \geq 1, \quad \textit{such that} \quad \alpha_n \neq 0 \qquad l_n(\alpha) \leq l_n(\kappa)$$

We will call the above inequalities the *circular inequalities.*

Note the following corollary, showing that, when the sequence α is finite, we can decide whether the circular inequalities are satisfied.

Corollary 1 *A finite sequence $\alpha = (\alpha_n)_{1\leq n\leq m}$ of nonnegative integers is the length distribution of a circular code over a weighted alphabet (A, λ) having length distribution $(\kappa_n)_{n\geq 1}$ if, and only if,*

$$\text{for} \quad 1 \leq n \leq m \quad \text{such that} \quad \alpha_n \neq 0, \qquad l_n(\alpha) \leq l_n(\kappa).$$

We can therefore decide whether a finite sequence is the length distribution of a circular code over a given weighted alphabet.

Remark 5 We can notice that verifying the circular inequalities till the rank m only in the case of a finite sequence $\alpha = (\alpha_n)_{1\leq n\leq m}$ is not necessarily obvious. Indeed, the shortest circular word having more than one circular factorization over a code which is not circular may be longer than the words of the code. For example, if $X = \{ab, cd, bc, da\}$, the shortest ambiguous circular word, $abcd$, is of length 4.

The proof is analogous to that of the Schützenberger theorem initially presented in [Sch65b], and later in [BP85] Theorem 3.6 p.343. The parallel comes from the fact that the properties of the conjugacy and X-conjugacy relations over a circular code X can be easily extended to a weighted alphabet.

Construction Let A be a weighted alphabet having length distribution $(\kappa_n)_{n\geq 1}$. We construct a sequence of prefix circular codes $(Y_i)_{i\geq 0}$ from the weigthed alphabet A. The circular code that we finally obtain hass, for all integers i, the same set of words of length smaller or equal to i as Y_i and contains exactly α_i words of length i. The construction of the codes $(Y_i)_{i\geq 0}$ is the following

$$Y_0 = A$$

and for all $i \in \mathbb{N}^*$, if $(y_{i,j})_{j\geq 1}$ is the length distribution of the code Y_i, the code Y_{i+1} is obtained by suppressing successively

$$\epsilon_i = y_{i-1,i} - \alpha_i = l_i(\kappa) - l_i(\alpha)$$

words of length i in the code Y_{i-1}: $x_{i,1}, x_{i,2}, \ldots, x_{i,\epsilon_i}$ and by constructing the auxiliary codes

$$X_{i,1} = x_{i,1}^*(Y_{i-1} \setminus x_{i,1}),$$

$$X_{i,2} = x_{i,2}^*(X_{i,1} \setminus x_{i,2}),$$

$$\ldots$$

$$Y_i = X_{i,\epsilon_i} = x_{i,\epsilon_i}^*(X_{i,\epsilon_i - 1} \setminus x_{i,\epsilon_i}).$$

Example 2 Let $k = 3$ and $A = \{a, b, c\}$. If $\alpha_1 = 1$, $\alpha_2 = 1$, $\alpha_3 = 7$, $\alpha_4 = 10$ and $\alpha_i = 0$ for $i \geq 5$, then $\epsilon_1 = l_1(3) - l_1(\alpha) = 2$, suppressing a, and then b, we get the code

$$Y_1 = \{c, \mid ab, ac, bc \mid, a^2b, a^2c, bab, bac, b^2c, a^3b, a^3c, ba^2b, ba^2c, b^3c, \ldots\}.$$

Since $\epsilon_2 = 2$, suppressing ab, and then ac, we get the code

$$\begin{aligned} Y_2 &= \{c, bc, \mid a^2b, a^2c, bab, bac, b^2c, abc, acc \mid, a^3b, a^3c, \\ &\quad ba^2b, ba^2c, b^2ab, b^2ac, b^3c, abac, abbc, acbc, \ldots\} \end{aligned}$$

Finally, since $\epsilon_3 = 0$ and $\epsilon_4 = 0$, $Y_2 = Y_3 = Y_4$. We obtain the following circular code Y having as length distribution $(1, 1, 7, 10)$

$$\begin{aligned} Y = Y_4 \cap A^4 &= \{c, bc, a^2b, a^2c, bab, bac, b^2c, abc, acc, \mid a^3b, a^3c, \\ &\quad ba^2b, ba^2c, b^2ab, b^2ac, b^3c, abac, abbc, acbc\} \end{aligned}$$

Remark 6

1. When α_i is zero, $\epsilon_i = y_{i-1,i}$ is nonnegative, and the inequality $l_i(\alpha) \leq l_i(\kappa)$ is satisfied. We can therefore ask whether there is another characterization showing directly this fact. Theorem 5 gives a positive answer to this question.

2. When A is a usual k-letter alphabet, the inequalities

$$\forall i \geq 1, \quad l_i(\alpha) \leq l_i(k) \tag{6}$$

are stronger than the Kraft-McMillan inequality. Moreover, the latter can be obtained directly from (6). Indeed, as $s_n = \sum_{d/n} dl_d$, we deduce from (6) that

$$\forall n \geq 1, \quad s_n(\alpha) \leq s_n(k) = k^n.$$

Furthermore denoting by $\alpha_n^{(*)}$ the numbers of words of length n in Y^*, we have:

$$\forall n \geq 1, \qquad \alpha_n^{(*)} \leq s_n(\alpha).$$

Thus, the convergence radius of the series $f_{Y^*}(z) = 1/(1 - f_Y(z))$ is greater than the one of f_{A^*}, that is, $1/k$. Therefore $\sum_{n \geq 1} \alpha_n k^{-n} \leq 1$.

3. The previous construction does not allow us to establish the rationality of Y even when we assume that the sequence α is $\mathbb{N}$-rational.

We now recall how is defined the operation of *composition* of codes. Let $X \subset A^*$ and $Z \subset B^*$ be two codes over the alphabets A and B respectively together with a bijection $\phi : B \to X$. Then the set $Y = \phi(Z)$ obtained by coding the letters of Z by the corresponding words of X is again a code over the alphabet A. We say that Y is obtained by composition of Z and X and we denote $Y = Z \circ X$.

The following result can be deduced from the two previous ones by making use of a coding morphism and by noticing that the code obtained by composition of two circular codes is again a circular code.

Corollary 2 *Let X be a circular code over a (usual) alphabet A and $\kappa = (\kappa_n)_{n\geq 1}$ be its length distribution. Then a sequence of nonnegative integers $\alpha = (\alpha_n)_{n\geq 1}$ is the length distribution of a circular code $Y \subset X^+$ if, and only if,*

$$\forall n \geq 1 \quad \textit{such that} \quad \alpha_n \neq 0, \qquad l_n(\alpha) \leq l_n(\kappa).$$

Proof : Since X is a code, for some alphabet B there exists an injective morphism $\phi : B^* \to A^*$ such that $X = \phi(B)$. We define length λ in B^* in the following way

$$\forall x \in B, \quad \lambda(x) = |\phi(x)|,$$

where $|\phi(x)|$ denotes the "usual" length of $\phi(x)$ in A^*, *i.e.*, the number of occurrences of letters of A in $\phi(x)$. We thus obtain a weighted alphabet, whose length distribution coincides with that of X.

From Theorem 2 (p.8), we can therefore construct a circular code Z over B^* having α as length distribution. Since the X and Z are circular codes and the code obtained by composition of two circular codes is again a circular code, $Y = Z \circ X = \phi(Z)$ is a circular code over A and consequently over X^* because moreover $Y^* \subset X^*$.

Finally, as direct consequence of the definition of λ, Y has α as length distribution over the alphabet A. $\square$

4 A new formulation of the circular inequalities

In this section, we show (Theorem 5) that length distributions $(\alpha_n)_{n\geq 1}$ of circular codes over an arbitrary weighted alphabet can be characterized by inequalities directly concerning the coefficients α_n instead of the numbers $l_n(\alpha)$ of conjugacy classes of primtive words of length n. This result underscores the fact that we can decide whether a finite sequence is the length distribution of a circular code over a given weighted alphabet.

Results

Let A be a weighted alphabet having length distribution $(\kappa_n)_{n\geq 1}$. From the previous construction (p.9), we deduce an expression with nonnegative coefficients of $\alpha_n + \epsilon_n$ in terms of $(\kappa_i)_{i\leq n}$, of $(\alpha_i)_{i<n}$ and of $(\epsilon_i)_{i<n}$ (where $\epsilon_i = l_i(\kappa) - l_i(\alpha)$).

Proposition 5 *Let (A, λ) be a weighted alphabet having length distribution $\kappa = (\kappa_i)_{i\geq 1}$ and let $\alpha = (\alpha_i)_{1\geq 1}$ be the length distribution of a circular code over A. Then for all positive integers n, the sum $\epsilon_n + \alpha_n$, where $\epsilon_i = l_i(\kappa) - l_i(\alpha)$, is a polynomial Q_n with nonnegative coefficients in the variables $(\epsilon_i)_{i<n}$, $(\alpha_i)_{i<n}$ and $(\kappa_i)_{i\leq n}$.*

The following theorem is then a direct consequence of Proposition 5 and of the previous theorem characterizing length distributions of circular codes (Theorem 2).

Theorem 3 *Let (A, λ) be a weighted alphabet having length distribution $\kappa = (\kappa_i)_{i\geq 1}$ and let $\alpha = (\alpha_i)_{i\geq 1}$ be a sequence of nonnegative integers. Then, there exists a sequence $(Q_n)_{n\geq 1}$ of polynomials Q_n in the variables $(\kappa_i)_{i\leq n}, (\alpha_i)_{i<n}$ and $(\epsilon_i)_{i<n}$ with nonnegative coefficients, such that the following conditions are equivalent*

1. the sequence $\alpha = (\alpha_i)_{i\geq 1}$ *is the length distribution of a circular code over* A

2. $\forall n \geq 1, \qquad \alpha_n \leq Q_n$.

Direct computation of the polynomials $Q_n (= \alpha_n + \epsilon_n)$ gives, for the first values of n

$$Q_1 = \alpha_1 + \epsilon_1 = \kappa_1, \qquad Q_2 = \epsilon_1 \alpha_1 + \binom{\epsilon_1}{2} + \kappa_2,$$

$$Q_3 = \left(\binom{\epsilon_1 + 1}{2} + \epsilon_2\right)\alpha_1 + 2\binom{\epsilon_1 + 1}{3} + \epsilon_1\kappa_2 + \kappa_3$$

$$Q_4 = \left(\binom{\epsilon_1 + 2}{3} + \epsilon_3\right)\alpha_1 + \epsilon_2\alpha_2 + 3\binom{\epsilon_1 + 2}{4} + \binom{\epsilon_2}{2} + \binom{\epsilon_1 + 1}{2}\kappa_2 + \epsilon_1\kappa_3 + \kappa_4.$$

Now we give a combinatorial proof of Proposition 5. By construction (p.9), the number of words of length n in Y_{n-1} is equal to $\alpha_n + \epsilon_n$. Studying the way in which these words are generated, we give another expression of this sum corresponding to the announced polynomial Q_n.

Case of a k-letter usual alphabet In order to count the number of words of length n in the code Y_{n-1}, we distinguish two distincts kinds among them:

1. the products of words that have been suppressed one of the construction steps:

$$w = w_p w_{p-1} \cdots w_1$$

where the w_i's $(1 \leq i \leq p)$ are words that have been previously suppressed.

2. such words concatenated with a word itself belonging to the code Y_{n-1}

$$w = w_p w_{p-1} \cdots w_1 w_0, \qquad w_0 \in Y_{n-1}$$

where the w_i's $(1 \leq i \leq p)$ are words that have been suppressed.

As the construction is made by successive suppressions of words of nondecreasing length, the order in which the words have been suppressed induces a total order on the set of these words (the first suppressed word being the smallest) and conditions the structure of the concatenations. Therefore, the products, which are prefixes of a word in the code or themselves elements of the code, can be factorized in a unique way in a nonincreasing sequence (for this order) of previously suppressed words:

$$\forall i, \qquad |w_{i+1}| \geq |w_i|.$$

Nevertheless, when the product itself belongs to the code, the last element is greater than the one that precedes it: if $w = w_p w_{p-1} \cdots w_1$ then $|w_1| = |w_2|$ and w_1 has been suppressed after w_2.

1. Words of the first kind:
More precisely, the words of length n of the form $w_p w_{p-1} \cdots w_1$ where the w_i's $(1 \leq i \leq p)$ are previously suppressed words (and have, consequently, a length strictly smaller than n) satisfy the following conditions on lengths:

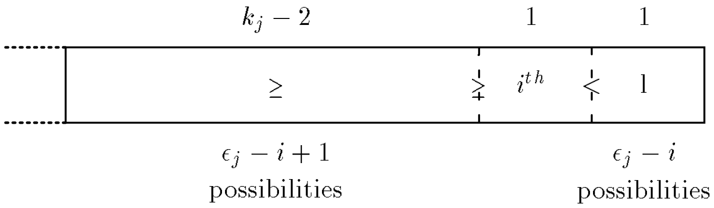

Figure 2: Suffixes formed from the ϵ_j suppressed words of length j

- either $|w_{i+1}| = |w_i|$,
- or $|w_{i+1}| > |w_i|$ and $|w_{i+1}| < |w_i w_{i-1} \ldots w_1|$.
  Otherwise,(*i.e.* if $|w_{i+1}| \geq |w_i w_{i-1} \ldots w_1|$), $W = w_i w_{i-1} \ldots w_1$ is itself either a word of the code and w is then a word of the second kind that we distinguished, or a word that we have suppressed and then w has to be factorized in the shorter way: $w = w_p \cdots w_{i+1} W$ with $|w_{i+1}| = |W|$.

We first compute the number of products of k_j elements of same length that may appear in such a word. We may observe that the elements occurring in a such product can be repeated, but their order is determined in a unique way.

If the associated factor is not in a suffix position (*i.e.*, $|w_1| \neq j$), this number is equal to the number of the k_j-combinations with repetitions of the set of the ϵ_j suppressed elements of length j, that is $\binom{\epsilon_j + k_j - 1}{k_j} = P_{k_j}$.

When this product is in a suffix position (*i.e.*, $\lambda(w_1) = j$), k_j is necessarily greater than or equal to 2 and the last element of the product is greater (for the order of suppression) than the previous one (Figure 2).

Let i be the rank of the next to last element of the product, then the number of associated prefixes having length $k_j - 2$ is equal to the number of the $(k_j - 2)$-combinations with repetitions of an $(\epsilon_j - i + 1)$-element set, that is $\binom{\epsilon_j + k_j - i - 2}{k_j - 2}$ and the number of values over which the last letter ranges is equal to $\epsilon_j - i$. The number of products of k_j elements having same length in a suffix position is thus equal to

$$T_{k_j} = \sum_{i=1}^{\epsilon_j - 1} \binom{\epsilon_j + k_j - i - 2}{k_j - 2} (\epsilon_j - i) = (k_j - 1) \sum_{i=1}^{\epsilon_j - 1} \binom{\epsilon_j + k_j - i - 2}{k_j - 1} = (k_j - 1) \binom{\epsilon_j + k_j - 2}{k_j}.$$

Consequently, the number of words having length n in Y_{n-1} and obtained by concatenations of suppressed words is

$$R_n = \sum_{\mathcal{P}} \prod_{j_1 < j_2 < \ldots < j_p} P_{k_{j_p}} \ldots P_{k_{j_2}} T_{k_{j_1}} \quad \text{with} \quad \begin{cases} T_{k_j} = (k_j - 1) \binom{\epsilon_j + k_j - 2}{k_j} & \text{and} \quad k_j > 1 \\ P_{k_j} = \binom{\epsilon_j + k_j - 1}{k_j} \end{cases}$$

where the sum is taken over the set $\mathcal{P}$ of p-uples satisfying the two following conditions (the second one expressing the growth of the lengths of the factors)

$$\sum_l j_l k_{j_l} = n \qquad \text{and} \qquad \forall p, \quad j_p < \sum_{l=1}^{p-1} j_l k_{j_l}.$$

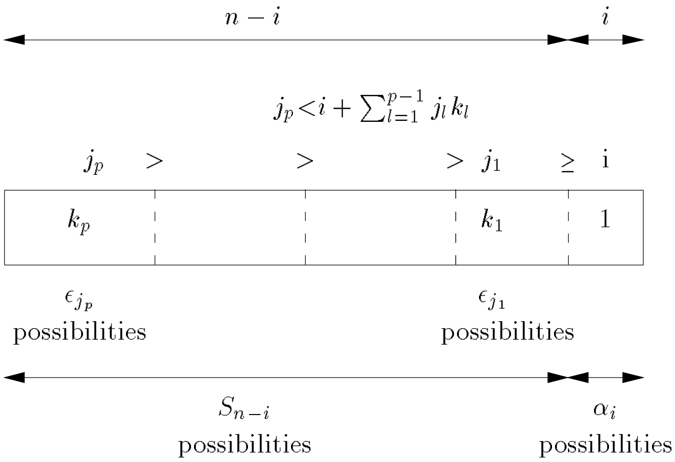

Figure 3: Words formed from the suppressed words and from a word of the code

2. Words of the second kind:
The other words of length n in the code Y_{n-1} can be written $w = w_p w_{p-1} \cdots w_1 w_0$ where the w_i's ($1 \leq i \leq p$) are suppressed words and w_0 is a word of the code $Y_{n-1} \setminus A$ which had been generated before the suppression of the word w_1. The w_i's satisfy the following conditions on the lengths (Figure 3):

- $\forall i \geq 0, \quad |w_{i+1}| \geq |w_i|$,
- $\forall i \geq 1, \quad |w_{i+1}| < |w_i w_{i-1} \ldots w_1 w_0|$.
  Because, if $|w_{i+1}| \geq |w_i w_{i-1} \ldots w_1 w_0|$, then $W = w_i w_{i-1} \ldots w_1 w_0$ is itself, either a word of the code and then w has to be factorized in the shorter way: $w_p \cdots w_{i+1} W$, or a word that we have suppressed and w is then a word of the first kind that we distinguished.

Let S_{n-i} be the number of products $w_p w_{p-1} \cdots w_1$ of length $n-i$ that can be concatenated with a word w_0 of Y_{n-1} of length i, then

$$S_{n-i} = \sum_{\mathcal{S}} \prod_{i \leq j_1 < j_2 < \ldots < j_p} P_{k_{j_p}} \ldots P_{k_{j_2}} P_{k_{j_1}} \quad \text{with} \quad P_{k_j} = \binom{\epsilon_j + k_j - 1}{k_j},$$

(the P_{k_j}'s denote the number of k_j-combinations of an ϵ_j-element set with repetitions) where the sum is taken over the set $\mathcal{S}$ of the p-uples satisfying the following conditions (the condition on j_k expressing the relations between the lengths of the factors)

$$\sum_l j_l k_{j_l} = n - i, \qquad i \leq j_1 \quad \text{and} \qquad \forall p \geq 2, \quad j_p < i + \sum_{l=1}^{p-1} j_l k_{j_l}.$$

At the end, the sum $\alpha_n + \epsilon_n$, equal to the number of words of length n in the code Y_{n-1}, can be written

$$\boxed{\alpha_1 + \epsilon_1 = k, \qquad \forall n \geq 2, \qquad \alpha_n + \epsilon_n = \sum_{i=1}^{\lfloor n/2 \rfloor} S_{n-i} \alpha_i + R_n.}$$

where $\lfloor x \rfloor = \max\{n/n \leq x, n \in \mathbb{N}\}$.

Example 2 (continued) Let $k = 3$ and $A = \{a, b, c\}$. We have

$$Y_1 = \{c, \mid \underline{ab}, ac, bc \mid, a^2b, a^2c, bab, bac, b^2c, a^3b, a^3c, ba^2b, ba^2c, b^3c, \ldots\}$$

and

$$\alpha_2 + \epsilon_2 = P_{1_1}\alpha_1 + T_2 = \binom{\epsilon_1}{1}\alpha_1 + \binom{\epsilon_1}{2} = 3.$$

$$\begin{aligned} Y_2 \quad &= \quad \{c, bc, \mid \underline{a^2b}, a^2c, \underline{bab}, bac, b^2c, abc, acc \mid, a^3b, a^3c, \\ &\qquad ba^2b, ba^2c, b^2ab, b^2ac, b^3c, abac, abbc, acbc, \ldots\} \end{aligned}$$

and

$$\alpha_3 + \epsilon_3 = (P_{2_1} + P_{1_2})\alpha_1 + T_{3_1} = \left(\binom{\epsilon_1 + 1}{2} + \epsilon_2\right)\alpha_1 + 2\binom{\epsilon_1 + 1}{3} = 7.$$

$$\begin{aligned} Y_3 = Y_2 \quad &= \quad \{c, bc, a^2b, a^2c, bab, bac, b^2c, abc, acc, \mid \underline{a^3b}, a^3c, \\ &\qquad \underline{ba^2b}, ba^2c, \underline{b^2ab}, b^2ac, b^3c, \underline{abac}, abbc, acbc \mid, \ldots\} \end{aligned}$$

and

$$\alpha_4 + \epsilon_4 = \left(\binom{\epsilon_1 + 2}{3} + \epsilon_3\right)\alpha_1 + \epsilon_2\alpha_2 + 3\binom{\epsilon_1 + 2}{4} + \binom{\epsilon_2}{2} = 10.$$

The products of suppressed (underlined) words
- for $n = 2$: ab,
- for $n = 3$: a^2b, bab,
- for $n = 4$: $a^3b, ba^2b, b^2ab, abac$.

The other words can be decomposed in products of suppressed words and of a word belonging itself to the code, this last is then in a suffix position (for example, for $n = 4$: $a^3.c$, $b.a^2.c$, $b^2.a.c$, $ab.bc$, $ac.bc$).

Case of a weighted alphabet Let $(\kappa_n)_{n \geq 1}$ be the length distribution of the weighted alphabet (A, λ). The number of words of length n in the code Y_{n-1} is equal, by construction, to $\alpha_n + \epsilon_n$. Their description is almost identical to the previous one, but since the alphabet is weighted, it contains words of length strictly greater than 1. This leads to distinguish a third kind of words.

3. Third kind of words:
The words of length n of the code Y_{n-1} can also be written, in this case, as $w_p w_{p-1} \cdots w_1 w_0$ where the w_i's $(1 \leq i \leq p)$ are suppressed words and w_0 belongs to the alphabet. Moreover, the κ_i words of length i of the alphabet being counted among the ϵ_i words that we suppress or the α_i belonging to the code, the lengths of the factors satisfy (Figure 4)

- $\forall i \geq 1, \quad \lambda(w_{i+1}) < \lambda(w_i)$

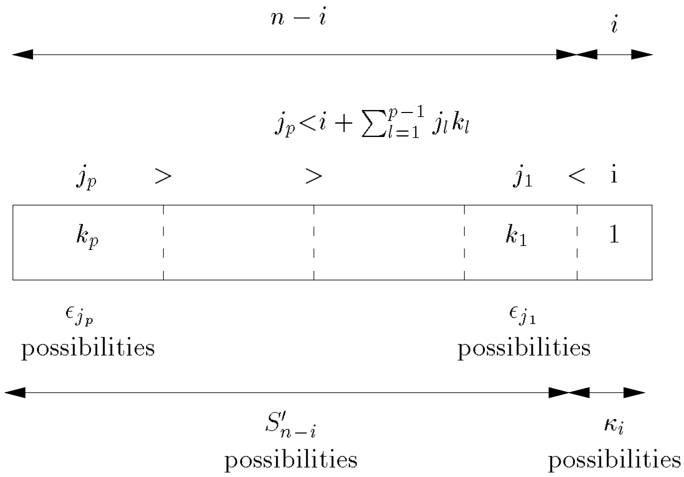

Figure 4: Words formed from the suppressed words and from a word of the alphabet

- and $\forall i \geq 0, \quad \lambda(w_{i+1}) < \lambda(w_i \cdots w_1 w_0)$.
  Because, if $\lambda(w_{i+1}) \geq \lambda(w_i \cdots w_1 w_0)$, then $w_i w_{i-1} \ldots w_1 w_0$ is itself either a word of the code, or a word that we suppress and $w_p w_{p-1} \ldots w_1 w_0$ is then a word of one of the two other kinds that we had studied before.

Then, by using the computation made for the prefixes of the words, the number of which is given by the coefficient S_{n-i}, and by considering the relations on the lengths of the factors, we obtain

$$S'_{n-i} = \sum_{\mathcal{S}'} \prod_{j_1<j_2<\ldots<j_p} P_{k_{j_p}} \ldots P_{k_{j_2}} P_{k_{j_1}} \quad \text{with} \quad P_{k_j} = \binom{\epsilon_j + k_j - 1}{k_j} \quad \text{and} \quad S'_0 = 1,$$

where the sum is taken over the set $\mathcal{S}'$ of p-uples such that

$$\sum_l j_l k_{j_l} = n - i, \qquad i > j_1 \quad \text{and} \qquad \forall p \geq 2, \quad j_p < i + \sum_{l=1}^{p-1} j_l k_{j_l}.$$

At the end, the sum $\alpha_n + \epsilon_n$, equal to the number of words of length n in the code Y_{n-1}, can be written

$$\alpha_1 + \epsilon_1 = \kappa_1,$$

$$\boxed{\forall n \geq 2, \qquad \alpha_n + \epsilon_n = \sum_{i=1}^{\lfloor n/2 \rfloor} S_{n-i}\alpha_i + R_n + \sum_{i=\lceil n/2 \rceil}^{n} S'_{n-i}\kappa_i.}$$

where $\lfloor x \rfloor = \max\{n/n \leq x, n \in \mathbb{N}\}$ and $\lceil x \rceil = \min\{n/n \geq x, n \in \mathbb{N}\}$.

5 Maximality and completeness

In this section, we establish a necessary and sufficient condition for a sequence of nonnegative integers to be the length distribution of a maximal circular code over a finite alphabet A. This criterion focuses on the asymptotic behavior of the number of conjugacy classes of primitive words.

We recall that a code X is said to be *maximal* over the alphabet A if it is not properly contained in any other code over the same alphabet, *i.e.*,

$$\text{if} \quad X \subset X', \quad X' \quad \text{code} \qquad \Rightarrow \qquad X = X'.$$

Remark 7 A maximal circular code over the alphabet A is necessarily infinite, except if $X = A$. Indeed, if X is finite, then for any letter $a \in A$ there exists an integer $n \geq 1$ such that $a^n \in X$. Since X is circular, $n = 1$ and consequently $X = A$.

A subset X of A^* is said to be *complete* if all words of A^* are factors of a word of X^*, and *thin* if, moreover, there exists a word of A^* that is not a factor of any word of X^*. A code is *synchronous* if there exist two words x and y of X^* such that for $u, v \in A^*$

$$uxyv \in X^* \quad \Rightarrow \quad ux, vy \in X^*.$$

A thin circular code is synchronous.

We also recall two properties that will be useful in the following. A thin code is maximal as circular code if, and only if, it is complete. A thin code X is complete and synchronous if, and only if,

$$\exists x, y \in X^* \quad \text{such that} \quad xA^*y \subset X^*.$$

We now compare the number of classes of primitive words in A^*, when A is a k-letter alphabet, and the number of classes of primitive words in X^* when X is a circular code.

Let $l_n(k)$ be the number of classes of primitive words of length n over a k-letter (usual) alphabet, then

$$\lim_{n\to\infty} \frac{nl_n(k)}{k^n} = 1.$$

In the case of circular codes over a k-letter alphabet, one can show that the asymptotic behavior of the number of classes of primitive words in $X^* \cap A^n$ has the following property.

Property 1 *Let X be a thin circular code over a k-letter alphabet A and let $(\alpha_n)_{n\geq 1}$ be its length distribution, then the three following limits are equal*

$$\limsup_{n\to\infty} \frac{s_n(\alpha)}{k^n} = \limsup_{n\to\infty} \frac{nl_n(\alpha)}{k^n} = \limsup_{n\to\infty} \frac{l_n(\alpha)}{l_n(k)} = \delta_X.$$

Their common value δ_X is called the *circular density* of X in the monoid A^*.

Proof : Since $s_n(\alpha) = \sum_{d/n} dl_d(\alpha)$ and

$$\forall d/n, \quad d \neq n, \qquad dl_d(\alpha) \leq \frac{n}{2}k^{n/2},$$

we get

$$\limsup_{n\to\infty} \frac{nl_n(\alpha)}{k^n} \leq \limsup_{n\to\infty} \frac{s_n(\alpha)}{k^n} \leq \limsup_{n\to\infty} \left(\frac{nl_n(\alpha)}{k^n} + \frac{n^2}{4}k^{-n/2}\right) = \limsup_{n\to\infty} \frac{nl_n(\alpha)}{k^n}.$$

Moreover, as $\lim_{n\to\infty} \frac{nl_n(k)}{k^n} = 1$, we obtain

$$\limsup_{n\to\infty} \frac{s_n(\alpha)}{k^n} = \limsup_{n\to\infty} \frac{nl_n(\alpha)}{k^n} = \limsup_{n\to\infty} \frac{l_n(\alpha)}{l_n(k)},$$

concluding the proof. □

The result below can be informally stated in this way: a circular code X is maximal over A if and only if the asymptotic behavior of the number of conjugacy classes of primitive words of X^* and of A^* are of the same order of magnitude. More formally:

Theorem 4 *Let X be a thin circular code having length distribution $(\alpha_n)_{n\geq 1}$, then the code X is maximal over A where A is a k-letter alphabet if, and only if, its circular density δ_X is positive.*

Proof : If the thin code X is maximal, it is complete. Since it is thin and circular, it is also synchronous, hence there exist two words x and y of X^* such that

$$xA^*y \subset X^*.$$

We denote by l the sum of the lengths of words x and y, let $l = |x| + |y|$. Then we have

$$\forall n > 0, \quad xA^n y \subset \left(X^* \cap A^{n+l}\right) \quad \text{and} \quad s_{n+l}(\alpha) \geq k^n.$$

We deduce that

$$\forall n > 0, \quad \frac{s_{n+l}(\alpha)}{k^{n+l}} \geq \frac{1}{k^l}.$$

Consequently, if the code is maximal

$$\exists l \in \mathbb{N}^*, \qquad \limsup_{n\to\infty} \frac{s_n(\alpha)}{k^n} > \frac{1}{k^l} > 0,$$

showing that this last condition is necessary.

On the other hand, if the code X is not maximal, it is not complete, and therefore, there exists a word y in A^* of length n_0 that is not a factor of any word in X^*. Let w be a word of length $n \geq n_0$, whose conjugate belongs to X^* then y is not a factor of w. Indeed, if y was a factor of a word $w = uv$ having a conjugate $t = vu$ in X^*, then y would be a factor of $t^2 = vuvu$. So, the longest prefix of w having as length a multiple of n_0 can be encoded over $A^{n_0} \setminus \{y\}$. We deduce that

$$\exists n_0 = |y| \quad \text{such that} \quad \forall n \geq n_0, \qquad s_n \leq (k^{n_0} - 1)^{\lfloor n/n_0 \rfloor} k^{(n \bmod n_0)},$$

where $\lfloor x \rfloor = \max\{n/n \leq x, n \in \mathbb{N}\}$.

Then we obtain

$$\limsup_{n\to\infty} \frac{s_n(\alpha)}{k^n} \leq \lim_{n\to\infty} (1 - \frac{1}{k^{n_0}})^{\lfloor n/n_0 \rfloor} = 0.$$

Consequently, if

$$\limsup_{n\to\infty} \frac{s_n(\alpha)}{k^n} \neq 0,$$

then there exists a maximal code over a k-letter alphabet having length distribution α. □

Remark 8 Notice that the general method (see [ER85] or [BP85] p.62) for completing codes also allows us to complete a circular code conserving this property.

More precisely, let X be a circular code and let $y \in A^*$ be an unbordered word (*i.e.*, no proper nonempty left factor of y is also a right factor) which is not a factor of any word of X^* and let $U = A^* \setminus X^* \setminus A^*yA^*$. Then $Y = X \cup y(Uy)^*$ is a complete circular code (see [Bas96]).

An important problem remaining open concerns the characterization of the length distributions of rational circular codes.

References

[AM95] D. Arquès and Christian Michel. A possible code in the genetic code. In E. W. Mayr and C. Puech, editors, *STACS 95*, number 900 in LNCS, pages 640–651. Springer, 1995.

[Bas96] F. Bassino. *Séries rationnelles et distributions de longueurs.* Thèse, Université de Marne-La-Vallée, 1996.

[BL96] V. Bruyère and M. Latteux. Variable-length maximal codes. In *ICALP 96*, number 1099 in LNCS, pages 24–47. Springer, 1996.

[BP85] J. Berstel and D. Perrin. *Theory of codes.* Pure and Applied Mathematics. Academic Press, Inc., 1985.

[Csi69] I. Csiszar. Simple proofs of some theorems on noiseless channels. *Inform. and Control*, 14:185–198, 1969.

[ER85] A. Ehrenfeucht and G. Rozenberg. Each regular code is included in a regular maximal code. *RAIRO Inform. Theor. Appl.*, 20:89–96, 1985.

[Gau96] H. Gaudier. Le lemme des signes. preprint, 1996.

[GG65] S. W. Golomb and B. Gordon. Codes with bounded synchronisation delay. *Inform. and Contr.*, 8:355–372, 1965.

[GGW58] S. W. Golomb, B. Gordon, and L. R. Welch. Comma-free codes. *Canadian J. Math.*, 10:202–209, 1958.

[Kra62] R. M. Krause. Channels which transmit letters of unequal duration. *Inform. and Control*, 5:13–24, 1962.

[Kro87] D. Krob. Codes limités et factorisations finies du monoïde libre. *RAIRO Inform. Theor. Appl.*, 21:437–467, 1987.

[MR83] N. Metropolis and G.-C. Rota. Witt vectors and the algebra of necklaces. *Adv. Math.*, 50:95–125, 1983.

[Reu88] C. Reutenauer. Mots circulaires et polynômes irréductibles. *Ann. Sc. Math. Québec*, 12(2):275–285, 1988.

[Sch65a] M. P. Schützenberger. On a factorization of free monoids. *Proc. Amer. Math. Soc.*, 16:21–24, 1965.

[Sch65b] M. P. Schützenberger. Sur une question concernant certains sous-monoïdes libres. *C. R. A. S. Paris*, 261:2419–2420, 1965.

[Sha48] C. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423,623–656, 1948.

[Sta86] R. P. Stanley. *Enumerative Combinatorics*, volume I of *Mathematics Series.* Wadsworth & Brooks/Cole, 1986.

[Vie78] G. Viennot. *Algèbres de Lie libres et monoïdes libres.* Number 691 in Lect. Notes in Math. Springer-Verlag, 1978.